

Faculty of Electrical and Electronic Engineering Technology



MUHAMAD AKMAL BIN RAZALI

Bachelor of Electronics Engineering Technology (Telecommunications) with Honours

2022

WIFI-BASED INDOOR LOCALIZATION UTILIZING MACHINE LEARNING TECHNIQUE

MUHAMAD AKMAL BIN RAZALI

A project report submitted in partial fulfillment of the requirements for the degree of Bachelor of Electronics Engineering Technology (Telecommunications)



UNIVERSITI TEKNIKAL MALAYSIA MELAKA

DECLARATION

I declare that this project report entitled "WIFI-BASED INDOOR LOCALIZATION UTILIZING MACHINE LEARNING TECHNIQUE" is the result of my own research except as cited in the references. The project report has not been accepted for any degree and is not concurrently submitted in candidature of any other degree.

Signature	Attand .
Student Nam	e : MUHAMAD AKMAL BIN RAZALI
Date	: 23 Feb. 23 اونيۇىرسىتى تىكنىكل مليسيا ملاك
ī	JNIVERSITI TEKNIKAL MALAYSIA MELAKA

APPROVAL

I approve that this Bachelor Degree Project 1 (PSM1) report entitled "WIFI-BASED INDOOR LOCALIZATION UTILIZING MACHINE LEARNING TECHNIQUE" is sufficient for submission.



DEDICATION

Alhamdulillah, praise to the Almighty Allah S.W.T

This thesis is dedicated to:

My beloved Parents,

MR RAZALI and MRS SAMSIAH

My supportive friends,



Ts. ZAHARIAH BINTI MANAP

,

ABSTRACT

Position estimation using global positioning systems (GPS) is not reliable in indoor environment due to weak signal penetration and the complex nature of indoor setting which causes severe signal attenuation. This project will explore the potential of utilising available wireless signal in indoor environment which is WiFi. We will implement fingerprinting technique to predict the position of mobile devices in a confined indoor space. The methodology involves two stages which are training stage and testing stage. In training stage, sufficient number of data will be measured in a study area. The data will be the received signal strength (RSS) measured by a mobile phone at predetermined reference points (RPs). The measured data will be trained using machine learning technique in Matlab platform to produce an indoor localization prediction model. In the testing stage, the produced model will be used to predict the location of mobile devices based on measured RSS. This project is expected to compare the accuracy produced by several machine learning techniques, and identify the best prediction model to be used in WiFi-based indoor localization.

ABSTRAK

Anggaran kedudukan menggunakan sistem penentududukan global (GPS) tidak boleh dipercayai dalam persekitaran dalaman disebabkan oleh penembusan isyarat yang lemah dan sifat tetapan dalaman yang kompleks yang menyebabkan pengecilan isyarat yang teruk. Projek ini akan meneroka potensi menggunakan isyarat wayarles yang tersedia dalam persekitaran dalaman iaitu WiFi. Kami akan melaksanakan teknik cap jari untuk meramalkan kedudukan peranti mudah alih dalam ruang tertutup terkurung. Metodologi melibatkan dua peringkat iaitu peringkat latihan dan peringkat ujian. Dalam peringkat latihan, bilangan data yang mencukupi akan diukur di kawasan kajian. Data tersebut akan menjadi kekuatan isyarat diterima (RSS) yang diukur oleh telefon mudah alih pada titik rujukan (RPs) yang telah ditetapkan. Data yang diukur akan dilatih menggunakan teknik pembelajaran mesin dalam platform Matlab untuk menghasilkan model ramalan penyetempatan dalaman. Dalam peringkat ujian, model yang dihasilkan akan digunakan untuk meramalkan lokasi peranti mudah alih berdasarkan RSS diukur serta-merta. Projek ini dijangka membandingkan ketepatan yang dihasilkan oleh beberapa teknik pembelajaran mesin, dan mengenal pasti model ramalan terbaik untuk digunakan dalam penyetempatan dalaman berasaskan WiFi.

ACKNOWLEDGEMENTS

First and foremost, I would like to express my gratitude to my supervisor, Madam Zahariah Binti Manap for her precious guidance, words of wisdom and patient throughout this project.

I am also indebted to Universiti Teknikal Malaysia Melaka (UTeM) which enables me to accomplish the project. Not forgetting my fellow colleague, Faris and Zahir for the willingness of sharing their thoughts and ideas regarding the project.

Finally, I would like to thank all the staffs at the Universiti Teknikal Malaysia Melaka (UTeM), fellow colleagues and classmates, the Faculty members, as well as other individuals who are not listed here for being co-operative and helpful.



TABLE OF CONTENTS

	PAGI
DECLARATION	
APPROVAL	
DEDICATIONS	
ABSTRACT	i
ABSTRAK	ii
ACKNOWLEDGEMENTS	iii
TABLE OF CONTENTS	i
LIST OF TABLES	iii
LIST OF FIGURES	iv
LIST OF SYMBOLS	v
LIST OF ABBREVIATIONS	vi
LIST OF APPENDICES	vii
اونيوم سيتي TITLE يونيوم سيتي APPENDIX ملاك APPENDIX	
CHAPTER INVERSIT OF THE ALAYSIA MELAKA	9
1.1 Background 1.2 Problem Statement	9 10
1.3 Project Objective	10
1.4 Scope of Project	11
CHAPTER 2 LITERATURE REVIEW	13
2.1 Introduction	13
2.2 Overview of Localization System 2.2.1 Technologies for Indoor Localization	15
2.2.1.1 Ultrasonic Based Localization	15
2.2.1.2 Infrared	18
2.2.1.3 ZigBee	18
2.2.1.4 Bluetooth	19
2.2.1.5 Wi-Fi Technologies	20
2.3 Summary	23
CHAPTER 3 METHODOLOGY	24
3.1 Introduction	24
3.2 Methodology	24

3.3	Developments of Fingerprint Database	26
3.4	Development of Indoor localization Prediction Model	29
3.5	Performance Evaluation	30
	3.5.1 Experimental Setup	31
	3.5.2 Regression Model	33
3.6	Summary	38
CHAI	PTER 4 RESULT AND ANALYSIS	39
4.1	Introduction	39
4.2	Data	39
	4.2.1 Data Collection	39
	4.2.2 Estimate Position and RMSE	41
	4.2.3 Training Regression Model	42
	4.2.4 Data Collections of RMSE Models Regression Learner	43
	4.2.5 Testing Data	46
	4.2.6 The Cumulative Distribution for Indoor Localization	49
4.3	Summary	50
CHAI	PTER 5 CONCLUSION AND FUTURE WORKS	52
5.1	Conclusion	52
5.2	Future Works	53
REFE	CRENCES	54
APPE		56
	*AIN0	
	اونيۆمرسيتي تيڪنيڪل مليسيا ملاك	
	UNIVERSITI TEKNIKAL MALAYSIA MELAKA	

LIST OF TABLES

TABLE

TITLE

PAGE

Table 2-1: Comparison between technologies used Indoor Localization14Table 4-1: True position and estimate positionError! Bookmark not defined.Table 4-2-: Root Mean Square Error (Estimate position)Error! Bookmark not defined.Table 4-3: Root Mean Square Error (Simulation position)Error! Bookmark not defined.



LIST OF FIGURES

FIGURE	TITLE	PAGE	
Figure 2.1:	Ultrasonic range finder detecting object	16	
Figure 3.1	Fingerprint flowchart	25	



LIST OF SYMBOLS



LIST OF ABBREVIATIONS



LIST OF APPENDICES



PAGE



CHAPTER 1

INTRODUCTION

1.1 Background

There are numerous outdoor positioning systems available, including GPS, A-GPS, Galileo, and others. However, they are only available outside and cannot give the accuracy required for indoor location. GPS, for instance, has a maximum accuracy of 5 metres. An accurate precision that is inappropriate for interior spaces. Outdoor accuracy can be accomplished using the Global Positioning System (GPS), but indoor performance is limited due to insufficient signal levels. Indoor localization has been proposed using a variety of technologies, including magnetic fields, Bluetooth, and WiFi. WiFi has the best availability and accuracy of all of them. Because of this, accurate localization is available in practically any setting and on almost any device. However, WiFi-based localization remains a challenge.

The WiFi-based location accuracy is calculated by measuring the distances between the mobile phone and the accessible Access Points (APs). The Received Signal Strength Indicator (RSSI) is used to figure out how far WiFi-based location estimate has to go. Many strategies for localization have been developed based on the RSSI, which are split into two broad categories signal propagation models and fingerprinting methodologies. Because of the random propagation effects, using the path loss propagation model in indoor environments makes distance estimate impossible, necessitating the use of multiple models in different regions. In interior situations, the fingerprint technique is extensively used due to its resilience and inexpensive cost. This project will look into the possibility of using WiFi as a wireless signal in an indoor context. In a constrained indoor setting, we will use a fingerprinting technique to forecast the location of mobile devices. The training stage and the testing stage are the two steps of the process. A sufficient number of data will be measured in a study region during the training stage. The data will be received signal strength (RSS), which will be measured by a mobile phone at predefined reference points (RPs). To create an indoor localization prediction model, the measured data will be trained using machine learning techniques in the Matlab platform. In this project, we only consider line-of-sight (LoS) conditions where there is no blockage between the mobile phone (target) and the WiFi access points (APs). The methods are based on regression models which provide a smooth position prediction.

1.2 Problem Statement

Outdoor positioning accuracy can be accomplished using the GPS that can provide an positioning error of about 5 metres. However, indoor positioning systems can't rely on GPS due to high signal penetration loss that reduce the positioning performance. Indoor localization has been proposed using a variety of technologies, including magnetic fields, Bluetooth, and WiFi. WiFi has the best availability and accuracy of all of them. Because of this, accurate localization is available in practically any setting and on almost any device. However, WiFi-based localization remains a challenge mainly due to rapid fluctuation in RSS. This problem causes error in measuring the distances between the mobile phone and the accessible APs.

Many strategies for localization have been developed based on the RSSI, which are split into two broad categories signal propagation models and fingerprinting methodologies. Because of the random propagation effects, using the path loss propagation model in indoor environments makes distance estimate impossible, necessitating the use of multiple models in different regions. Therefore, in this project, we will implement fingerprinting technique which is more resilience and inexpensive cost. We will trained the measured data to obtain the best localization prediction model with the highest accuracy.

1.3 Project Objective

The aim of this project is to:

- To develop a fingerprints reference database through manual measurement of received signal strength (RSS) by using mobile application.
- 2. To develop an indoor localization model for specific scenario by using machine learning techniques.

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

3. To evaluate the performance of the developed indoor localization model.

1.4 Scope of Project

The scope of this project are as follows:

 The proposed project were tested in a Double Storey House but only conducted on 1st Floor.

- 2. There were four available WiFi Access Points (APs) located in the house .
- 3. A floor map was created based on the 1st Floor of the House that being tested and the WiFi Access Points (APs) located in the floor.
- 4. An android application was used on a mobile device. The application was used to collect RSS measurements from the available Access Points (APs).
- 5. The radio map is created using the Receive Signal Strength (RSS) readings that have been obtained.
- 6. We consider line-of-sight (LOS) scenario where the mobile phone (mobile device) and WiFi access points (Aps) are not blocked.
 Use the second scenario where the mobile phone (mobile device) and WiFi access points (Aps) are not blocked.

EKNIKAL MALAYSIA MELAKA

a)

UNIVERSIT

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

Due to weak signal penetration and the complex structure of indoor settings, which causes severe signal attenuation, position estimation using GPS is not reliable in an indoor context. This project will look into the possibility of using the available WiFi signal in an interior context. In a constrained indoor setting, we will use fingerprinting to forecast the location of mobile devices. The training and testing stages are both included in the system.

A sufficient number of data will be measured in a study region during the training stage. The data will be received signal strength (RSS), which will be measured by a mobile phone at predefined reference points (RPs). To create an indoor localization prediction model, the measured data will be trained using machine learning techniques in the Matlab platform. The developed model will be utilised in the testing stage to forecast the position of mobile devices based on real-time RSS measurements. The goal of this research is to assess the accuracy of several machine learning approaches and determine the optimal prediction model for WiFi-based indoor localisation.

2.2 Overview of Localization System

Indoor localization system technologies could be a forward-thinking invention, similar to how the Global Positioning System (GPS) revolutionised outdoor navigation. According to the authors (Michael Brown and James Pinchin, 2013), no single indoor localization technology available today can give an absolute solution for every indoor localization case. However, a variety of technologies are being offered as a means of developing and improving the indoor localization system.

A detail survey on indoor localization system can be found in [1]. The authors described the technologies, techniques and algorithms used for indoor localization system. The most common and widely used technologies as mentioned in the articles are Satellite, Ultra-wideband (UWB), Wi-Fi, ZigBee and Bluetooth. Table 2.1 illustrates the comparison between technologies used in IPS which done by authors in [1].

Technology	Cost	Coverage	Advantages	Disadvantages
Satellite	ALAYSIA	Floor level	Low power	-
S.		2	consumption	
Ultrasonic	Medium-	Room level	-Good accuracy	-Interference
Based	High		-No offect of	-Cost for hardware
600	_ ==		multipath	
	alun .		-Cheap	
Infrared	Medium	Room level	-Cheap	-Sunlight interference
	** **	(few meters)	-No effect of	-Short-range
UNIV	/ERSITI T	EKNIKAL M	multipathA MELAK	-Cost for hardware
			-Low power	
			consumption	
Visible Light	Medium	Floor level	No interfering	Expensive construction
Wi-fi	Low	Floor level	-Good Accuracy	-RF interference with
		(around 35)	-Low cost	operating at 2.4 GHz
			-Signals can penetrate	-Fingerprinting requires
			walls	a huge effort
Zighee	Medium	Floor level	-Low cost	-Require special
Ziguee	wicului		Low power	-require special
			-Low power	equipment
			consumption	

Table 2-1: Comparison between technologies used Indoor Localization

Bluetooth	Low-	Floor level	-Good accuracu	-RF interference	
	Medium	(around 10)	-No need for	-Limited coverage and	
			additional	mobility	
			infrastructure		
			-Low power		
			consumption		
Inertial	Low	Floor level	Cheap	-Accumulative errors	
Magnetic Based	Low	Floor level	Cheap	Requires mapping	

Based on all technologies above, Wi-Fi does not demand any additional hardware or infrastructure. Wi-Fi infrastructure is already available in nearly all locations, offering a basic level of positioning capability without additional investment. Manually deployed hardware solutions can be expensive and time intensive, especially if you have many sites that require the capacity. As a result, other technologies need a bigger financial and time investment in order to establish an effective indoor localization environment.

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

2.2.1 Technologies for Indoor Localization

Interior localization systems are classed according to how they determine their most common technology for locating indoor surroundings. Radio frequency (RF), inertial, audible and non-audible sound, light-based, and vision-based technologies are commonly

used.

2.2.1.1 Ultrasonic Based Localization

Ultrasonic is a well-known indoor positioning technology. Ultrasound offers a few features that make it easy to use in a positioning system, such as a moderate speed of propagation, minimal wall penetration, and low component costs. The ultrasonic localization based system as a system that uses ultrasound with sound frequencies higher than audible range (more than 20 kHz). In comparison to other indoor positioning systems, the ultrasonic system has an active and passive system. The characteristics of the ultrasound pulse are fascinating to be implemented in IPS, because it has good precision, which is normally in millimetres. The ultrasonic range finder was used to detect an object in Figure 2.1.



Figure 2.1: Ultrasonic range finder detecting object

Because of its low-power penetration losses through indoor walls, low-cost components, and compatibility with handled equipment, ultrasonic localisation is preferred within short ranges. Localization errors, on the other hand, are introduced by a variety of variables, including numerous reflections from surfaces and synchronisation issues between communicating nodes. Complex signal processing algorithms plague ultrasonic localization systems. Ultrasonic localization systems may withstand low time synchronisation better than Ultra-Wideband (UWB) systems. The localization median error was 15 cm. Before localization, the location of ultrasonic nodes should be determined.

The coordinates of the object are calculated analytically, and a matrix equation based on distance measurements, sometimes known as the Least Squares algorithm, is constructed. The current coordinate of the moving item is obtained utilising a number of measurements between the number of stationary transceivers and the moving object based on the analytical description of the moving object. A pseudo pyramid of measurement of transceivers and moving objects is built using the technique. To detect the target, five ultrasonic HX7TR produced ultrasound pulses.

The physical amount of signal is obtained using the ToA technique, and the distance estimating strategy is based on shape coordinating. The system based on ultrasonic location uses a narrow band transmission signal. This is because the narrow band allows the device to function in the same frequency bandwidth, resulting in a more precise signal and a more steady communication range. To acquire the signal envelope, the receiver identifies the transmitter's signal peak. The downward zero-intersection in the first subsidiary that exceeds the inclination limit is used to detect this peak . The time interval between the received signal and the reference signal is measured using the cross correlation approach. The arrival of probable echoes following the LOS signal does not interfere with the positioning system's measurement. According to the findings, the ultrasonic signal's starting point has an impact on the positioning system's accuracy, robustness, and multipath problem.

2.2.1.2 Infrared

In Line of Sight (LOS) circumstances, where handsets already have sensors such as photodiodes, infrared (IR) devices are used. IR is known for its ease of use, light weight, small size, and resilience to interference (unlike RF systems). However, it is affected by fluorescent light and sunshine, and the hardware of these systems is expensive and timeconsuming to maintain. IR systems are made up of infrared emitting devices (such as LEDs) and infrared sensors (e.g., photodiode). The target wears an infrared emitting device that generates a signal with a unique identity; after the sensors detect the IR signal, the target's location is determined. Commercial IR-based technologies, such as the Active Badge, are an example. The Active Badge system provides a means of locating individuals within a building by determining the location of their Active Badge. This small device worn by personnel transmits a unique infra-red signal every 10 seconds. PIR sensors are used to detect heat radiation changes released by people and animals, and their accuracy is measured in centimetres; nevertheless, they are susceptible to environmental changes. They have a wide range of applications that it is nearly unthinkable to imagine everyday modern life without EKNIKAL them. These sensors are common in PIR-based motion detectors. Examples are security alarms, door opening, automatic lighting switches, vending machines, lift lobbies.

2.2.1.3 ZigBee

ZigBee is an IEEE 802.15.4 standard-based specification. It operates on the 868 MHz band in Europe, 915 MHz in the United States and Australia, and 2.4 GHz in the rest of the world. In a wireless mesh network, ZigBee is utilised for long-distance communication between devices. In comparison to WiFi standards, it is inexpensive, has a low data transfer rate, and has a short latency time. The RSS method is used to estimate the distance between

two or more ZigBee sensor units in this technology. The scanning of access points (APs) via the WiFi interface consumes a lot of electricity. To mitigate this effect, the ZigBee interface is used to collect Wi-Fi signals in an energy-efficient indoor localization system known as ZIL.

2.2.1.4 Bluetooth

Bluetooth (IEEE 802.15.1) is a technology that allows devices to communicate wirelessly across short distances. Bluetooth uses radio waves with frequencies ranging from 2.402 GHz to 2.480 GHz, similar to Wi-Fi. If you're looking for cost-effectiveness, low transmission power, long battery life, secure and efficient communications, and easily accessible solutions, look no further. Bluetooth Low Energy (BLE) is a new Bluetooth version that has a range of 70–100 metres and provides 24 Mbps with greater power efficiency. As a result, Bluetooth is not ideal for large-area localisation. In the training phase, neural networks (NN) are trained using received signal strength values and their corresponding coordinates; once learned, NN may be utilised to detect user location using online RSS measurements. The smartphone can be used for localization within airports, train stations, large markets, malls, and restaurants, where the area map is given to the smartphone and then localization is conducted via BLE.

The Bluetooth Special Interest Group created BLE technology, which is a wireless personal area network technology (SIG). According to the authors (Faragher and Harle, 2015), BLE technology competes with Wi-Fi technology. The author explains that because the BLE device operates in the license-free 2.4 GHz band, it can share the same indoor propagation characteristics as Wi-Fi transceivers. Nowadays, more indoor positioning systems are using BLE-based systems; this is due to the widespread usage of BLE in smartphones and its low power consumption, which allows stationary emitters to operate for months or years on batteries. Furthermore, when an object is in close proximity to a BLE beacon, the BLE system can establish its location. The new BLE system is supported by the majority of smart devices, allowing it to overcome the limitations of classic Bluetooth, such as extended scan times.

Furthermore, in comparison to other publications using BLE technology, the other articles focused more on examining the accuracy of BLE indoor placement with specific localization algorithms. They suggested a mix of channel separate polynomial regression model (PRM), channel separate fingerprinting (FP), outlier detection, and extended Kalman filtering for smart-phone based indoor localisation using BLE beacons (EKF). To assess the target's location and distance between BLE beacons, the fingerprinting (FP) and polynomial regression model (PRM) algorithms are used. Separate and aggregate location estimate strategies based on FP and PRM are compared to determine the accuracy distance for each strategy.

2.2.1.5 Wi-Fi Technologies

The popular wireless networking technology is known as Wi-Fi. IEEE 802.11b, IEEE 802.11 g, and IEEE802.11n use the 2.5 GHz RF spectrum, while IEEE 802.11a uses the 5 GHz range. WiFi hotspots that give whole-building coverage as a network access point have already been disseminated in the most vast indoor environments, such as a university or an office building. Personal computers, video gaming consoles, smartphones, digital cameras, tablet computers, and digital music players all use Wi-Fi technology. Wi-Fi infrastructure and user device costs can be very low, and Wi-Fi has grown from a reception range of around 100 metres to about 1 kilometre (km). Wi-Fi location based on fingerprinting RSS is also available (Received Signal Strength). Other RF localization systems, such as RFID, could be combined with Wi-Fi. Wi-Fi has a larger coverage area than Bluetooth and has a better throughput, making it more practical to use. RADAR, HORUS, COMPASS, HERECAST, and PlaceLab are examples of commercial Wi-Fi-based positioning systems [1].

Indoor localization systems based on fingerprint approaches can be divided into three categories: deterministic, probabilistic, and machine learning approaches. The wellknown WiFi RADAR system was the first RF-based localization system to use the WLAN fingerprint's deterministic technique. In the offline stage, the WiFi RADAR system saves RSSI fingerprints at grid points to construct the radio map. The K-nearest neighbourhood approach is used to determine the user's fingerprint's estimated location [2].

The probabilistic strategy is based on computing the probability of each grid point and applying Bayesian inference to estimate the user's position. In comparison to the WiFi RADAR system, which is based on the deterministic technique, the latter approach is used in the HORUS system, which yields better precision. The fundamental disadvantage of the probabilistic technique is that creating a distribution requires a large number of samples from APs. This lengthens the time it takes to create the radio map and necessitates a considerable amount of storage space.

Another probabilistic indoor location system is the COMPASS. By including a compass into the system, this method considers the attenuation induced by the human body. The COMPASS system builds numerous radio maps with various orientations (usually 45° or 90° each) in the offline stage. In the online stage, a digital compass is used to provide user direction, and only fingerprints with similar orientations are used to predict the user's location. The COMPASS system's main drawback is that it necessitates a huge amount of storage space in order to keep many radio maps with varied orientations.[2]

The method introduced employed the Wi-Fi Crowdsourced fingerprinting data set to gain the measurement of the object's position, which was different from earlier Wi-Fi based positioning systems. The experiment was also conducted in a large indoor location, such as a university building, rather than a tiny room to assess distance. The dataset is collected in the building using 21 android devices. The Gaussian likelihood location fingerprinting is used to estimate the preliminary position. There are a few challenges while collecting data, for example, the actual location of the access point in the building is unknown, and some of them affect the mobile access node since the user's Wi-Fi is in hotspot mode.

The system introduced in utilises the Wi-Fi Crowdsourced fingerprinting data set to gain the measurement of the object's position, as opposed to the traditional Wi-Fi based positioning system. The experiment was also conducted in a large indoor location, such as a university building, rather than a tiny room to assess distance. The dataset is collected in the building using 21 android devices. The Gaussian likelihood location fingerprinting is used to estimate the preliminary position. The authors claimed that there are a few challenges while collecting data, for example, the actual location of the access point in the building is unknown, and some of them affect the mobile access node since the user's Wi-Fi is in hotspot mode. The access point location is rough estimated using the weighted centroid estimator and data that has already been acquired. The weighted centroid is a simple approach for estimating indoor position that just considers the estimated access point. Aside from that, rank-based fingerprinting is utilised, in which the access point with the strongest signal is ranked first. This is due to the RSS rankings being bias and scaling insensitive. The mean positioning errors are in the 8-10m range, which is higher than earlier studies (3-6m). The system's drawback, according to the author, is that Wi-Fi fingerprinting is dependent on a variety of characteristics that are difficult to regulate in an experiment, such as access point (AP) distribution and architectural features.

2.3 Summary

According to the various papers that being reviewed, there are a wide range type of technologies for indoor localization being applied nowadays. From this literature review, it can be said that each of the technologies has different coverage, accuracies, application and limitation. However, the accuracy is the most importance aspects for the indoor localization system. Accuracy is refers to the variance between estimated position and actual position. Due to this aspect, most of the articles that being reviewed are more focusing in accuracy and positioning system improvement.



CHAPTER 3

METHODOLOGY

3.1 Introduction

The goal of this project is to create a fingerprint reference database using a mobile application to manually measure incoming signal strength (RSS). This chapter outlines the entire work process, methodology, and techniques used in the development of the proposed wifi-indoor localisation. Essentially, this project requires two items: Access Points (AP) and Mobile Devices. The following section will provide an overview of the work process, followed by a detailed explanation of the hardware and software development processes.

3.2 Methodology

The flowchart in Figure 3.1 depicts the project's process flow. The study begins with a survey of the literature on wifi indoor localization technologies and method Many research topics are discussed in articles and conference papers connected to this topic. It is found in the literature that Wifi Indoor localization attracts the majority of researchers' attention in the subject of indoor localization reaches. This is due to the fact that GPS is ineffective for inside tracking and localization due to restricted satellite coverage. Indoor localization systems are important for determining the location of objects in enclosed spaces. Wifi technology is one of the most reasonable and extensively utilised technologies for indoor location, according to my understanding gained from the survey. Therefore, we will use WiFi technology to provide a solution to indoor localization systems.



Figure 3.1 Fingerprint flowchart

The system design process is divided into two parts: Training Phase and Testing Phase. In the training phase, access points (APs) and mobile devices are involve based on figure 3.2. In the testing phase, the software and the algorithm employed within it are detailed. The project involves detecting the wifi signal from the access point and receiving it by mobile devices at various coordinates. Following that, a large amount of data is gathered in order to analyse the system's accuracy.



Figure 3.2 Device used (a) Mobile Device

3.3 Developments of Fingerprint Database

The Receive Signal Strength (RSS) collected from each grid points on the map is used for the data. Cabinets, tables, and other partitions in the experimental setting all have a line-of-sight (LOS) effect on signal transmission. We chose an experimental location with six Access Pints (APs), each of which was evenly spaced. The six APs will send Wi-Fi signals, while apps on mobile devices will collect data based on the RSS. Six APs from a few coordinates in the room make up the data. Based on the table 3-1, the dataset will be like :

T 11	A 1	C 1	C	D
Table	3-1	Sample	of	Dataset

<rss< th=""><th>AP</th><th>AP</th><th> AP</th><th>Position</th></rss<>	AP	AP	 AP	Position
>	1	2	n	

Rss1	-65	-90	•••	-78	(10.0,12.
					0)
Rss2	-70	-88		-73	(7.3,10.0
)
Rss3	-85	-80		-69	(2.6,7.9)
		6.4X			···×
Rssn	-90	-75		-65	(7.6,8.3)





Figure 3.3 Training flowchart

To perform WiFi localization, mobile device with WiFi 802.11 a/b/g/n/ac will be used in this test. Mobile devices use radio waves to broadcast and receive signals. The distinction is that mobile device operate on separate frequencies. The 824-MHz through 894-MHz frequency ranges are used by cell phones. 2.4 GHz is used by WiFi phones that implement the 802.11b or 802.11g protocols. Mobile device that support the 802.11a standard operate at a frequency of 5 GHz. For receivers, Cisco Aironet 1130 G Series will
be used. It offers a total throughput of 108 Mbps to client devices that comply with the 802.11a, 802.11b, and 802.11g wireless LAN standards. The Cisco Aironet 1130AG Series contains built-in diversity antennas that provide excellent coverage for offices and similar settings. WiFi Analyzer app will be used to measure the Wireless Signal Strength which is measured in dBm (decibel milliwatts), expresse in negative values.



The signal strength is expressed in decibels (dBm) (0 to -100). This is the decibel (dB) power ratio of the measured power when compared to one milliwatt. The stronger the signal, the closer the value is to 0.

3.4 Development of Indoor localization Prediction Model

To develop an indoor localization model for specific scenario by using machine learning techniques, we proposed Regression Learner in Matlab platform. Regression models describe the connection between one or more predictor (input) variables and a response (output) variable. Fit linear, generalised linear, and nonlinear regression models, including stepwise and mixed-effects models, with Statistics and Machine Learning ToolboxTM. You may use a model to forecast or simulate responses, check model fit using hypothesis tests, or show diagnostics, residuals, and interaction effects using plots.

The RSS vector is collected by the mobile device, and the ideal target position is computed by comparing the stored fingerprint and received vector. The radio map refers to the entire collection of RSS fingerprints for the indoor environment.



For the time being, Cisco Room that will be use are not prepared to collect data. So, any regressions model cannot be shown for now until enough data collected.

3.5 Performance Evaluation

In terms of apps that will be use which is Wifi Analyzer are functioning. Based on the figure 3.6, it shows all the access points that available also with the name and the MAC address.



To evaluate the performance of the proposed project in real environments, this UNIVERSITITEKNIKAL MALAYSIA MELAKA

3.5.1

paper gather the data of receive signal strength (RSS) from the experimental environment as shown in Figure 3.7. The localization were made with 92 points based on the coordinates plotted in the floor plan. Using the mobile device , all results are complete collected by walking from point 1 to point 92 along the path. The collected data are set to 2.4GHz and 5 GHz because the routers can operate in two different frequencies. For each point, the sampling process is executed three times. The experiments are carried out on the Xiaomi Redmi Note 11 comes with 6GB internal and 2GB virtual Ram. CPU is Snapdragon 680 with Octa-core and maximum frequency of 2.40 GHz. The Regression Learner is a mathematic model from Matlab are used to train regression models including linear regression models, regression trees, Gaussian process regression models, support vector machines, kernel approximation, ensembles of regression trees, and neural network regression models as in Figure 3.8. In phase 1 will be Training Phase which is estimated around 80% of data collected will be train and also there Test Phase and estimated around 20% of data collected will train in the Machine Learning Matlab.



UNIVERSITI TEKNIKAL MALAYSIA MELAKA Figure 3.7 Floor plan with Point Coordinates

And a second sec		115
New Control PCA	LINEAR REGRESSION MODELS	- 12
Session - Selection Optimizer		
FILE OPTIONS Models	Linear Interactions Robust Stepwise	-
Sort by Model Number * 1111	Linear Linear Linear	J.
🔄 1 Tree	All Linear	
Last change. Fine free		
	Linear SVM Quadratic Cubic SVM	
	Medium Coarse All SVMs Optimizable Gaussian _ Gaussian _ SVM	
	GET STARTED ± ‡	
	All Quick-To- All Train	
	REGRESSION TREES 🚖 🙄	
	Fine Tree Medium Tree Coarse Tree All Trees	
	Optimizable Tree	
	GAUSSIAN PROCESS REGRESSION MODELS 🚊 🍹	
	Rational Squared Matem 5/2 Exponential Cuedratic Exponential Optimizable Models GPR	
	KERNEL APPROXIMATION REGRESSION MODELS 🚊 🍹	
	SVM Kernel Least All Kernels	
	ENSEMBLES OF TREES	
	Boosted Bogged All Optimizable Trees Trees Ensembles	
14		- ins

Figure 3.8 Models in Regression Learner

3.5.2 Regression Model

For this project, I need to export my collected data which is originally from Microsoft Excel into the workspace to use the model with new data based on Figure 3.8 into Matlab. In Microsoft Excel, I collect the data that consists of :

- Point Location
- Point Coordinates (m)
- RSSI (dbm)
- 2.4 GHz Access Points (AP)
- 5 GHz Access Points (AP)

		В	с	E	F	G	н	. 1	L	к	LI	
10												
11	Point Location	Point Coor	dinates (m)				RSSI(-	-dBm)				
12					2.4	GH7			50	H.		
13					2.14	GIL					_	
14	-	x	Y	1x	2x	3x	Tag	1x	2x	3x	Tag	
15 16				-37	-31	-41	59	-41	.46	.55	EA	
17				-36	-45	-40	82	-52	-44	-59	83	
18	1	19.00	8.00	-62	-66	-66	06	-78	.71	-90	07	
19				-57	-52	-53	9E	-76	-68	-83	9F	
20		-		-33	-37	-42	F8	-51	-51	-50	FA	
21		19.00	-	-40	-40	-41	B2	-43	-40	-37	B3	
22	2		7.00	-58	-61	-63	06	-72	-74	-74	07	
23				53	-48	-49	9E	-56	-63	-64	9F	
24		10.00		-48	-40	-45	F8	-47	-49	-54	FA	
25			6.00	-39	-38	-39	B2	-53	-46	-54	B3	
26	3	19.00	6.00	-64	-61	-70	06	-81	-62	-81	07	
27				-53	-53	-43	9E	-74	-67	-84	9F	
28		19.00	5.00	-37	-34	-32	F8	-50	-50	-50	FA	
29	(2)			-43	-43	-42	B2	-49	-42	-41	B3	
30	4		5.00	-61	-60	-57	06	-66	-71	-74	07	
31				-53	-52	-53	9E	-57	-56	-56	9F	
32				-41	-43	-41	F8	-50	-47	-51	FA	
33	c	19.50	4.00	-42	-39	-39	82	-43	-48	-46	B 3	
34	5	10.50	4.00	-62	-63	-64	06	-67	-69	-70	07	
35				-49	-49	-50	9E	-67	-69	-67	9F	
36			-	-45	-36	-45	F8	-45	-56	-54	FA.	
37	6	18 50	2.50	-42	-40	-45	B2	-54	-52	-68	B3	
38	0	10.50	2.30	-61	-66	-62	06	-85	-77	-85	07	
39				-55	-58	-57	9E	-79	-71	-87	9F	

Figure 3.8 Collected in Microsoft Excel

When my data have already convert into Matlab workspace based on Figure 3.9, there will be 2 data which are Training Data and Test Data and it consists 11 important things which are :

- SSID1-2.4 GHz UNIVERSITI TEKNIKAL MALAYSIA MELAKA
- SSID2- 2.4 GHz
- SSID3- 2.4 GHz
- SSID4- 2.4 GHz
- SSID1- 5 GHz
- SSID2- 5 GHz
- SSID3- 5 GHz
- SSID4- 5 GHz
- X- Coordinate
- Y- Coordinate

Z	Editor - untitled3	11					🔏 Variab	oles - Testing					💿 ×
J	Testing X												
	55x11 table												
	1 SSID1-2.4G	2 SSID1-5G	3 SSID2-2.4G	4 SSID2-5G	5 SSID3-2.4G	6 SSID3-5G	7 SSID4-2.4G	8 SSID4-5G	9 X-Coordinate	10 Y-Coordinate	11 Point No.	12	13
1	-71	-79	-55	-77	-64	-90	-66	-89	7.5000	2	88		1
2	-42	-50	-41	-37	-63	-74	-49	-64	19	7	2		
3	-64	-78	-67	-82	-71	-91	-74	-100	6	3	90		
4	-69	-76	-62	-66	-70	-66	-70	-82	8	1	83		
5	-64	-72	-58	-69	-52	-77	-68	-90	8	8	66		
6	-54	-62	-55	-56	-68	-74	-74	-85	8.5000	-4	60		
7	-56	-69	-63	-71	-59	-68	-73	-83	6.5000	8	73		
8	-61	-72	-58	-64	-80	-80	-71	-86	7.5000	0.5000	84		
9	-48	-62	-40	-49	-51	-48	-55	-73	11	5	-49		
10	-73	-77	-74	-69	-57	-46	-62	-70	11	1	57		
r	-50	-58	-48	-58	-51	-68	-62	-82	9	4.5000	61		
12	-39	-54	-40	-57	-66	-80	-59	-83	16.5000	8	13		
13	-34	-52	-41	-43	-53	-65	-55	-59	16.5000	5	16		
14	-41	-50	-42	-43	-62	-67	-49	-67	18.5000	4	5		
15	-30	-48	-37	-39	-61	-65	-49	-67	17.5000	7	11		
16	-50	-65	-46	-50	-52	-50	-61	-72	12.5000	5	39		
17	-35	-48	-34	-38	-59	-65	-56	-67	17.5000	7	11		
18	-59	-56	-53	-52	-70	-61	-62	-75	10	5	46		
19	-50	-62	-53	-50	-68	-73	-63	-78	12.5000	3	37		
20	~36	-46	-47	-38	-61	-63	-60	-71	16.5000	7	14		
2-	-34	-51	-42	-40	-56	-67	-55	-58	16.5000	5	16		
22	-60	-77	-63	-71	-64	-74	-70	-87	8	5	74		





Figure 3.9 (a)Training Data and (b)Test Data

Automated Regression Model Training will be use to do the diagnostic measure such as models accuracy and plots. In this Regression Learner , multiple models will be run simultaneously. In figure 3.10, to access the Regression Learner app, click Regression Learner under the Machine Learning and Deep Learning group on the Apps menu. Select data from the workspace or from a file by clicking New Session (based in figure 3.11) in the file section of the Regression Learner tab. Include a response variable and the variables that will be used to predict it. Click the arrow to expand the list of regression models in the Models section. Choose Quick-To-Train for All. All of the model presets that are quick to fit are trained using this option. Select Train All by clicking Train All in the Train section. The Models pane displays a variety of model kinds. The best RMSE (Validation) score (figure 3.13) is highlighted in a box once the models have finished training.



Figure 3.12

Models	0
Sort by Model Number 🔹 🗼 🏌	
🔂 1 Tree	RMSE (Validation): 2.9263
Last change: Fine Tree	10/10 features
2.1 Linear Regression	RMSE (Validation): 6.7275
Last change: Linear	10/10 features
2.2 Linear Regression	RMSE (Validation): 7.5686
Last change: Interactions Linear	10/10 features
2.3 Linear Regression	RMSE (Validation): 6.7768
Last change: Robust Linear	10/10 features
2.4 Stepwise Linear Regression	RMSE (Validation): 5.8672
Last change: Stepwise Linear	10/10 features
3.1 Linear Regression	RMSE (Validation): 6.7275
Last change: Linear	10/10 features
3.2 Tree ALAYSIA	RMSE (Validation): 2.9263
Last change: Fine Tree	10/10 features
😭 3.3 Tree	RMSE (Validation): 4.3512
Last change: Medium Tree	10/10 features
😭 3.4 Tree	RMSE (Validation): 9.2488
Last change Coarse Tree	10/10 features
Figure 3 مليسيا ملاك	اونيۇم سىتى تى 1

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

3.6 Summary

Wifi Indoor Localization is described in this chapter utilising methods and procedures that have already been discussed. The essential component hardware for this project is an access point (AP) and a mobile device. The Matlab programme is used to acquire the data of prediction and Root Mean Square Error(RMSE) in simulation position. The RMSE for simulation position and estimate position are compared. Matlab is utilised to create the cumulative probability graph from this data. The localization system presented in this project tracks an object in the system region and provides the coordinates of the mobile device's position.



CHAPTER 4

RESULT AND ANALYSIS

4.1 Introduction

The objective of this project is to investigate the possibility of utilising the available WiFi signal in indoor environment. In the first stage, a sufficient number of RSS were measured in a study region by using a mobile phone. The measured data were extracted and prepared in a suitable format for machine learning application in Matlab. We use the Regression Learner to train and test the data. This chapter describes the measurement data and proposed model and analyse the findings.

4.2 Data

The experimental data for indoor localization is collected on the first floor of the Double Storey House. The project specifically focuses on RSS data, unlike channel state information (CSI), time of arrival (TOA), angle of arrival (AOA), which require additional hardware or installation.

4.2.1 Data Collection

RSS measurements are captured via the Wifi Analyzer App (figure 4.2) in mobile phone which provides an RSS sampling rate of about 1 samples/5 s. Data were manually recorded in Microsoft Excel as in Figure 4.1.

Point Location	Point Coordinates (m)		RSSI(-dBm)								
			2.4GHz				5GHz				
	x	Y	1 x	2x	3x	Tag	1x	2x	3х	Tag	
		8.00	-37	-31	-41	F8	-41	-46	-55	FA	
	19.00		-36	-45	-40	B2	-52	-44	-59	B3	
1			-62	-66	-66	06	-78	-71	-90	07	
			it	-52	-53	9E	-76	-68	-83	9F	
			-33	-37	-42	F8	-51	-51	-50	FA	
	10.00		-40	-40	-41	B2	-43	-40	-37	B3	
2	19.00	7.00	-58	-61	-63	06	-72	-74	-74	07	
			53	-48	-49	9E	-56	-63	-64	9F	

-59 -60 -57 F8 -57 -68 -72 FA -57 -56 -54 B2 -60 -62 -61 B3 91 7.00 3.00 -69 -68 -68 06 -71 -73 -73 07 -69 -70 -70 9F 9E -86 -87 -88 -57 -58 -56 -66 -57 -57 FA F8 -54 -56 -54 -54 -54 B3 -55 B2 92 8.00 3.00 -70 -69 -70 06 -62 -65 -65 07 -74 -68 -66 9E -100 -100 -100 9F

÷





Figure 4.2 Result from Wifi Analyzer App

Based on figure 4.2, 57dBm is better signal strength than -78dBm. The stronger the signal, the closer the value is to 0.

4.2.2 Estimate Position and RMSE

The key to ensuring that the project meets its goal is to use RMSE. The difference between the values predicted by a model or an estimator and the actual values observed is measured by the RMSE definition. The actual distance between Access Points (Ap) and the mobile device was used to determine how accurate Wifi-based Indoor Localization was. To understand the link between one or more predictor variables and a response variable, we can utilise the regression learner using Matlab. Calculating the root mean square error, a statistic that shows the average gap between the predicted values from the model and the actual values in the dataset, is one technique to judge how well a regression model fits a dataset. A given model can "fit" a dataset more accurately the lower the RMSE. The estimation error is calculated using the following formula (4.1).

$$Error = \sqrt{(X_{est} - X_true)^2 + (Y_{est} - Y_true)^2}$$
(4.1)

The performance of a regression model is frequently assessed using the statistic known as Root Mean Squared Error (RMSE). It calculates the typical difference between the expected and actual values. The square root of the mean squared error (MSE) between the forecasts and actual values is used to determine the RMSE. The model's predictions are stated to be more accurate and close to the actual values when the RMSE value is lower. The RMSE number is important when assessing a regression model.

4.2.3 Training Regression Model

A particular kind of machine learning technique called regression is used to forecast a continuous numerical value based on the properties of the input data. Finding the optimal collection of parameters that reduces the difference between the projected values and the actual values is the aim of training a regression model. A regression algorithm must be chosen, the data must be collected and preprocessed, the data must be divided into training and test sets, the model must be trained, its performance must be evaluated, the model must be adjusted, and the model must finally be deployed in a production environment. After that, the dataset is divided into two sets: one for training (Figure 4.4) the model and the other for assessing its efficiency. The appropriate algorithm is then chosen after that. Regression methods come in many varieties, including support vector regression, polynomial regression, and linear regression. The model's performance on the test set is assessed after it has been trained on the training set by contrasting the predicted values with the actual values. The model's correctness is measured using evaluation metrics including root mean squared error (RMSE), mean absolute error (MAE), R-squared value, and correlation coefficient, among others.



4.2.4 Data Collections of RMSE Models Regression Learner

Table 4-1 shows the RMSE for all models that have different values(refer appendices). Highest value of RMSE for the regression models is 25.36 at Support Vector Machines : Cubic SVM. While, lowest value of RMSE is Tree : Fine Tree which is 2.1545 (figure 4.5). The training RMSE (root mean squared error) value on a regression learner for WiFi localization based on RSS (received signal strength) can be used to evaluate the accuracy of the model. A lower RMSE value indicates that the model is making fewer errors and is more accurate in predicting the location of the device based on the RSS values.

Models (Regression Learner in Matlab)	RMSE
1. Linear Regression : Linear	6.5202

Table 4-1: RMSE for all Models

2. Linear Regression : Interactions Linear	7.333
3. Linear Regression : Robust Linear	6.6624
4. Linear Regression : Stepwise Linear	6.094
5. Tree : Fine Tree	2.1545
6. Tree : Medium Tree	3.9635
7. Tree : Coarse Tree	8.9623
8. SVM : Linear SVM	6.474
9. SVM : Quadratic SVM	8.4601
10. SVM : Cubic SVM	25.36
11. SVM : Fine Gaussian SVM	19.08
12. SVM : Medium Gaussian SVM	6.5115
13. SVM Coarse Gaussian SVM	7.5097
14. Ensemble : Bagged Trees	5.6878
15. Ensemble : Boosted Trees	3.2816
16. Gaussian Process : Squared Exponential	6.0842
17. Gaussian Process : Matern 5/2 GPR	15.4807AYSIA MELAKA
18. Gaussian Process : Exponential GPR	4.4187
19. Gaussian Process : Rational Quadratic GPR	4.8388
20. Neural Networks : Narrow	7.8632
21. Neural Networks : Medium	11.601
22. Neural Networks : Wide	7.401
23. Neural Networks : Bilayered	7.178
24. Neural Networks : Trilayered	6.3289
25. Kernel : SVM	16.351

Summary × Response Pic Model 1: Tree Status: Trained Training Results RMSE (Validation) 2.1 R-Squared (Validation) 1.4 MAE (Validation) 1.4 MAE (Validation) 1.4 Prediction speed -11 Training time 7.6 Test Results RMSE (Test) 1.1355 R-Squared (Test) 1.00 MSE (Test) 1.2893 MAE (Test) 0.8022 • Model Hyperparameters Preset: Fine Tree Minimum leaf size: 4 Surrogate decision split • Feature Selection: 10/10 • PCA: Disabled • Optimizer: Not applicab Fine Table 4-2: Training 4-2: Traini	545 19 1417 1424 600 obs/sec 5749 sec 15 15 16 16 19 19 10 10 10 10 10 10 10 10 10 10
RMSE (Validation)	2.1545
R-Squared (Validation)	اونىۋىمىسىتى تېك
MSE (Validation)	4.6417
UNIVERSITI TEKNIKAL	MALAYSIA MELAKA
MAE (Validation)	1.4424
Prediction Speed	1600obs/sec
Training Time	7.6749 sec



Figure 4.6

ALAYSIA

The decision tree model may be the best fit for the WiFi localization problem based on RSS values, according to the tree model in MATLAB that yields the lowest RMSE value of all the models employed in the regression learner. This indicates that, when compared to other models, the decision tree model makes the fewest errors in predicting the device's location based on the RSS readings based on response plot (figure 4.6).

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

4.2.5 Testing Data

Performance analysis and experimental results are introduced in this section. We first calculate the root mean square error (RMSE) of positioning using every model in the regression learner. Next, we choose the model with the lowest RMSE (root mean-squared error). The predicted coordination and location accuracy of these results is further examined. One set of data was used to train the model in regression learner, and the other set was used to assess how well it performed. 20% of the data (figure 4.7) that have been choose randomly are being tested using the model that have been trained. By comparing the predicted locations of the test set's devices with their actual locations, the test set can be used to measure how

well the model performs. In this way, you may assess the model's performance and decide whether or not the level of accuracy is acceptable for the application that it is intended for.



Table 4-3: RMSE for Estimated Location for X-Coordinate and Y-Coordinate UNIVERSITI TEKNIKAL MALAYSIA MELAKA

X_true	Y_true	X_estimated	Y_estimated	RMSE
18.5	4	18.50	4.00	0.00
10.5	7.5	10.50	7.50	0.00
8	7	8.00	7.00	0.00
17	4	17.00	4.00	0.00
10.5	7.5	10.50	7.50	0.00
19	7	19.00	7.00	0.00
9	7	9.00	7.00	0.00
17	4	17.00	4.00	0.00
11.5	2.5	11.50	2.50	0.00
18.5	4	18.50	4.00	0.00
12.5	5	12.50	5.00	0.00
14.5	3	14.50	3.00	0.00
15	7	15.00	7.00	0.00
15	7	15.00	7.00	0.00

T State Stat			1	· · · · · · · · · · · · · · · · · · ·
11	0.5	11.00	11.00 0.75	
11	0.5	11.00	0.75	0.25
7.5	0.5	7.75	0.75	0.35
9	1	9.25	0.75	0.35
9.5	0.5	9.25	0.75	0.35
7.5	2	7.00	2.00	0.50
7.5	2	7.00	2.00	0.50
8	4	8.00	4.50	0.50
8	5	8.00	4.50	0.50
10.5	0.5	10.50	1.00	0.50
10	4	10.00	4.50	0.50
10	5	10.00	4.50	0.50
16.5	8	16.50	7.50	0.50
7	7	6.75	7.50	0.56
12	6	11.25	6.00	0.75
12	6	11.25	6.00	0.75
8	3	7.00	3.00	1.00
6MAL	4	6.00	5.00	1.00
8	6	8.00	7.00	1.00
11.5	1.5	11.50	2.50	1.00
11.5	3.5	11.50	2.50	1.00
- 14.5	4	14.50	3.00	1.00
17.5	8	17.50	7.00	1.00
19/11	6	19.00	7.00	1.00
19	8	19.00	7.00	1.00
19	8.h.	19.00	7.00	1.00
7	1.3	6.00	1.50	1.02
13.5	CS 0.5 T	12.50	MAL1.00SIA	ME1.12 (A
14	7	13.50	6.00	1.12
14	7	13.50	6.00	1.12
19	5	18.5	4	1.12
19	5	18.5	4	1.12
17.5	3	17.00	4.00	1.12
13.5	4	12.50	3.00	1.41
7	6	5.50	6.00	1.50
7	6	5.50	6.00	1.50
15.5	4	15.50	2.50	1.50
18.5	2.5	18.5	4	1.50
12	8	10.50	7.50	1.58
12.5	7	12.50	5.00	2.00
13.5	8	13.50	6.00	2.00

Table 4-2 shows information such as X-Coordinate True and Y-Coordinate True which is randomly picked, Root Mean Squared Error (RMSE), X-Coordinate estimated and Y-Coordinate estimated. The Euclidean distance formula is sqrt((predicted x - true x)2 + (predicted y - true y)2), which is based on the 4.1. It is frequently used to assess the accuracy of predicted locations in WiFi localization and determines the distance between two points in a 2-dimensional space.



4.2.6 The Cumulative Distribution for Indoor Localization

As one technique to represent the distribution of random variables, The Cumulative Distribution Function (CDF) is used to compute the cumulative possibility and apply in data analysis. The RMSE of the actual and estimate coordinates is used to plot the cumulative distribution graph. The cumulative vs. RMSE graph is displayed in Figure 4.8 below. To make result analysis simpler, the dotted lines are employed.



Figure 4.8 Fingerprint flowchart

The cumulative percentage is directly related to the RMSE value, as seen in Figure 4.8. The CDF shows that 50% and 90% of the testing points are estimated with less than 30cm and 50cm of error, respectively. This 50% of probability is a central estimation of data where the median value of RMSE is located.

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

4.3 Summary

From this chapter, the analysis of data obtains from the Wifi Indoor Localization is the stronger the signal, the closer the value is to 0. Then, in order to achieve the best result for this WiFi Indoor Localization project, we will compute the Root Mean Squared Error (RMSE) and compare it to the values that the machine learning models will predict. The best model performance is demonstrated by Fine Tree Regression, which has the lowest RMSE (2.1555).



CHAPTER 5

CONCLUSION AND FUTURE WORKS

5.1 Conclusion

In this paper, Wi-Fi indoor localization system using Regression Learner as machine learning has been proposed. In the room situations, real experiments have been carried out using WiFi Analyzer application. In these circumstances, the proposed technique obtain the best localization performance with low cost and good accuracy. Using a regression learner based on Received Signal Strength (RSS) for WiFi indoor localization can be an effective way to predict the location of a device in an indoor environment. By training a model on RSS measurements collected at known locations, the model can be used to predict the location of a device based on its RSS measurements. However, the accuracy of the predictions can vary depending on the complexity of the indoor environment, the number and distribution of RSS measurements, and the quality of the data. Also, the dynamic changes in the indoor environment, such as the presence of people or furniture, which can influence RSS data, are likely to have an impact on the prediction' accuracy. The Fine Tree model produced the project's lowest root mean square error (RMSE) value. As seen by the model's RMSE score of 2.1555, it appears to be the most accurate and exact model tested for locating a device using measurements of its received signal strength (RSS). The CDF shows that 50% and 90% of the testing points are estimated with less than 30cm and 50cm of error, respectively

5.2 Future Works

For future works , hybrid networks should be add with Regression which is Convolutional Neural Network (CNN) and Long-Short Term Memory(LSTM). CNN and LSTM are type of artificial neural network, which is widely used for object recognition and classification. Both of the machine learning will increase the accuracy of indoor localization. Then, examine the ideal number and placement of access points for a certain environment in our future works due to the relevance of the number and locations of access points. 5GHz frequency also will be used in the future to make sure that the result will get the best accuracy.



REFERENCES

- [1] Koovimol, P., & Pattaramalai, S. (2021). Experimental machine learning for RSSI fingerprint in Indoor WiFi Localization. ECTI-CON 2021 2021 18th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology: Smart Electrical System and Technology, Proceedings, 1018–1021. https://doi.org/10.1109/ECTI-CON51831.2021.9454865
- [2] Chen, Z., Zou, H., Yang, J. F., Jiang, H., & Xie, L. (2020). WiFi Fingerprinting Indoor Localization Using Local Feature-Based Deep LSTM. IEEE Systems Journal, 14(2), 3001–3010. https://doi.org/10.1109/JSYST.2019.2918678
- [3] Hsieh, C. H., Chen, J. Y., & Nien, B. H. (2019). Deep Learning-Based Indoor Localization Using Received Signal Strength and Channel State Information. IEEE Access, 7, 33256–33267. https://doi.org/10.1109/ACCESS.2019.2903487
- [4] Zhang, M., Jia, J., Chen, J., Deng, Y., Wang, X., & Aghvami, A. H. (2021). Indoor Localization Fusing WiFi with Smartphone Inertial Sensors Using LSTM Networks.
 IEEE Internet Strof EK Things Journal, 14 8(17), 13608–13623. https://doi.org/10.1109/JIOT.2021.3067515
- [5] Aloui, N., Raoof, K., Bouallegue, A., Letourneur, S., & Zaibi, S. (2014). Performance evaluation of an acoustic indoor localization system based on a fingerprinting technique. In EURASIP Journal on Advances in Signal Processing (Issue 13). http://asp.eurasipjournals.com/content/2014/1/13
- [6] Babalola, O. P., & Balyan, V. (2021). Wifi fingerprinting indoor localization based on dynamic mode decomposition feature selection with hidden markov model. Sensors, 21(20). https://doi.org/10.3390/s21206778

[7]



APPENDICES



A. Regression Learner Models RMSE (Training Set)

Models	0
Sort by RMSE (Validation)	
😭 1 Tree	RMSE (Validation): 2.1545
Last change: Fine Tree	10/10 features
2.2 Tree	RMSE (Validation): 2.1545
Last change: Fine Tree	10/10 features
S Tree	RMSE (Validation): 2.1545
Last change: Fine Tree	10/10 features
6.2 Tree	RMSE (Validation): 2.1545
Last change: Fine Tree	10/10 features
7.5 Tree	RMSE (Validation): 2.1545
Last change: Fine Tree	10/10 features
2.14 Ensemble	RMSE (Validation): 3.2816
Last change: Boosted Trees	10/10 features
2.3 Tree ALAYSIA	RMSE (Validation): 3.9635
Last change: Medium Tree	10/10 features
6.3 Tree	RMSE (Validation): 3.9635
Last change: Medium Tree	10/10 features
7.6 Tree	RMSE (Validation): 3.9635
Last change: Medium Tree	10/10 features
7.18 Gaussian Process Regression	RMSE (Validation): 4.4187
Last change: Exponential GPR	10/10 features
7.19 Gaussian Process Regression	RMSE (Validation): 4,8388
Last change: Rational Quadratic GPR	10/10 features
7.17 Gaussian Process Regression IKAL M	RMSE (Validation): 5.4807
Last change: Matern 5/2 GPR	10/10 features
7.15 Ensemble	RMSE (Validation): 5.6878
Last change: Bagged Trees	10/10 features
7.16 Gaussian Process Regression	RMSE (Validation): 6.0842
Last change: Squared Exponential GPR	10/10 features
7.4 Stepwise Linear Regression	RMSE (Validation): 6.094
Last change: Stepwise Linear	10/10 features
7.24 Neural Network	RMSE (Validation): 6.3289
Last change: Trilayered Neural Network	10/10 features

Models	6)	
Sort by RMSE (Validation) V	1	1	
7.8 SVM	RMSE (Validation): 6.474		
Last change: Linear SVM	10/10 features		
7.12 SVM	RMSE (Validation): 6.5115	ĺ	
Last change: Medium Gaussian SVM	10/10 features		
2.1 Linear Regression	RMSE (Validation): 6.5202		
Last change: Linear	10/10 features		
6.1 Linear Regression	RMSE (Validation): 6.5202		
Last change: Linear	10/10 features		
7.1 Linear Regression	RMSE (Validation): 6.5202		
Last change: Linear	10/10 features		
7.3 Linear Regression	RMSE (Validation): 6.6624		
Last change: Robust Linear	10/10 features		
7.23 Neural Network	RMSE (Validation): 7.178		
Last change: Bilayered Neural Network	10/10 features		
7.2 Linear Regression	RMSE (Validation): 7.333		
Last change: Interactions Linear	10/10 features		
7.22 Neural Network	RMSE (Validation): 7.401		
Last change: Wide Neural Network	10/10 features		
⑦ 7.13 SVM	RMSE (Validation): 7.5097		
Last change: Coarse Gaussian SVM	10/10 features		
7.20 Neural Network	RMSE (Validation): 7.8632		
Last change: Narrow Neural Network	10/10 features		
⑦ 7.9 SVM ALL	RMSE (Validation): 8.4601		1. raig
Last change: Quadratic SVM 📑 🚽	10/10 features		05.5
	RMSE (Validation): 8,9623	1	
Last change: Coarse Tree	10/10 features		MELANA
6.4 Tree	RMSE (Validation): 8.9623		
Last change: Coarse Tree	10/10 features		
7.7 Tree	RMSE (Validation): 8.9623		
Last change: Coarse Tree	10/10 features		
7.26 Kernel	RMSE (Validation): 10.593		
Last change: Least Squares Regression Kernel	10/10 features		
7.21 Neural Network	RMSE (Validation): 11.601	1	
Last change: Medium Neural Network	10/10 features		
7.25 Kernel	RMSE (Validation): 16.351		
Last change: SVM Kernel	10/10 features		
7.11 SVM	RMSE (Validation): 19.08		
Last change: Fine Gaussian SVM	10/10 features		
7.10 SVM	RMSE (Validation): 25.36		
Last change: Cubic SVM	10/10 features	¥	

B. Exported Data Model to Predict Coordinate for Test Set



1	Variables - Testi	ng											⊙×
	Testing X												
	55x12 table												
	1 SSID1-2.4G	2 SSID1-5G	3 SSID2-2.4G	4 SSID2-5G	5 SSID3-2.4G	6 SSID3-5G	7 SSID4-2.4G	8 SSID4-5G	9 X-Coordinate	10 Y-Coordinate	11 Point No.	12 New Point.no	13
1	-71	-79	-55	-77	-64	-90	-66	-89	7.5000	2	88	89	^
2	-42	-50	-41	-37	-63	-74	-49	-64	19	7	2	. 2	
3	-64	-78	-67	-82	-71	-91	-74	-100	6	3	90	89	
4	-69	-76	-62	-66	-70	-66	-70	-82	8	1	83	84.2000	
5	-64	-72	-58	-69	-52	-77	-68	-90	8	8	66	67.1250	
6	- 54	-62	- 55	-56	-68	-74	-74	-85	8.5000	4	60	61.1429	
7	-56	-69	-63	-71	-59	-68	-73	-83	6.5000	8	73	71.8750	
8	-61	-72	-58	-64	-80	-80	-71	-86	7.5000	0.5000	84	84.2000	
9	-48	-62	-40	-49	-51	-48	-55	-73	11	5	49	49,6000	
10	-73	-77	-74	-69	-57	-46	-62	-70	11	1	57	57.5556	
11	-50	-58	-48	-58	-51	-68	-62	-82	9	4.5000	61	61.1429	
12	-39	-54	-40	-57	-66	-80	- 59	-83	16.5000	8	13	14.5714	
13	-34	-52	-41	-43	-53	-65	-55	-59	16.5000	5	16	14.5714	
14	-41	-50	-42	-43	-62	-67	-49	-67	18.5000	4	5	5	
15	-30	-48	-37	-39	-61	-65	-49	-67	17.5000	7	11	11	
16	-50	-65	-46	-50	-52	-50	-61	-72	12.5000	5	39	39	
17	-35	-48	-34	-38	-59	-65	-56	-67	17.5000	7	11	11	
18	- 59	-56	- 53	-52	-70	-61	-62	-75	10	5	46	47.1429	
19	- 50	-62	- 53	-50	-68	-73	-63	-78	12.5000	3	37	37	
20	-36	-46	-47	-38	-61	-63	-60	-71	16.5000	1	14	14.5714	~ 7
21	-34	-51	-42	-40	-56	-67	-55	-58	16.5000	5	16	14.5714	14
22	-60	-77	-63	-71	-64	-74	-70	-87	8	5	74	. 75	~
	<												>

UNIVERSITI TEKNIKAL MALAYSIA MELAKA



	/ariables - Testin	ng											\odot
ſ	Testing X					a set or							
	55x12 table				WAL	ATSIA							
1	1 🗸	2	3	4	5	6	7	8	9	10	11	12	13
	SSID1-2.4G	SSID1-5G	SSID2-2.4G	SSID2-5G	SSID3-2.4G	SSID3-5G	SSID4-2.4G	SSID4-5G	X-Coordinate	Y-Coordinate	Point No.	New Point.no	
	-63	-77	-61	-72	-64	-73	-73	-87	8	5	74	77,3333	_
	-69	-84	-69	-76	-55	-49	-60	-73	11	0.5000	56	5 57.5556	
	-58	-63	-54	-66	-66	-67	-75	-89	13.5000	0.5000	32	32.6000	
1	-32	- 50	-42	-41	-57	-74	-53	-56	19	5	4	4 5	
	-66	-68	-57	-55	-66	-73	-69	-88	7	1.3000	87	86.4000	
i.	-64	-77	-62	-75	-61	-82	-76	-100	6	4	76	77.3333	
1	-52	-66	-51	-58	-53	-58	-63	-80	11.5000	2.5000	52	51.2500	
	-52	-65	-44	-48	-51	- 50	-61	-73	12.5000	5	39	39	
	- 50	-58	-56	- 50	-53	- 55	-64	-77	14.5000	3	29	29.8571	
2	-36	-56	-40	-52	-66	-77	- 58	-71	18.5000	2.5000	6	5 5	
3	-37	-41	-36	-52	-62	-78	-57	-76	19	8	1	2	
1	-45	-54	-39	-54	-70	-81	-43	-84	19	6	3	3 2	
5	-39	-56	-39	-47	-40	- 50	-61	-75	12,5000		39	26	. 4
5	-68	-71	-66	- 54	-54	-49	-60	-73	10.5000	1.5000	59	57.5556	1.1
	-63	-81	-64	-68	-59	-66	-72	-83	9	8	65	63.8750	φч
3	-31	-46	-33	-26	-61	-71	-54	-61	15	7	22	21.3750	14.
1	-73	-77	-60	-66	-76	-75	-76	-88	7.5000	0.5000	84	84.2000	
)	-64	-100	-69	-82	-65	-89	-81	-100	6	5	77	77.3333	
	-38	-53	-34	-54	-52	-76	-58	-89	13.5000	8	24	4 26	AV
	-61	-70	-52	-61	- 58	-65	-71	-78	11.5000	2.5000	52	51.2500	
	-61	-77	-68	-78	- 55	-75	-74	-88	5.5000	6	70	69.4000	
1	-30	-46	-32	-25	-61	-69	-53	-61	15	7	22	21.3750	



					HIRS	MAL	N 8/4	HELPH					
esti	ng				<u> </u>		~	2			1	⊙×	
5					1		-						
					-								
	2 SSID1-5G	3 SSID2-2.4G	4 SSID2-5G	5 SSID3-2.4G	6 SSID3-5G	7 SSID4-2.4G	8 SSID4-5G	9 X-Coordinate	10 Y-Coordinate	11 Point No.	12 New Point.no	13	
-55	-64	-54	-61	-58	-52	-67	-75	11.5000	1.5000	53	54.1250	^	
47	-63	-51	-51	-63	-59	-68	-82	10	4	47	47.1429		
-68	-76	-61	-66	-71	-78	-71	-87	7.5000	0.5000	84	84.2000		
-36	-42	-47	-50	-62	-75	-57	-73	16.5000	8	13	14.5714		
50	-43	- <mark>4</mark> 6	-45	-65	-74	-56	-69	17.5000	3	7	8.1250	1	
-50	-66	-56	-58	-54	-63	-71	-84	14.5000		31	29.8571	Auto	P, ~ ~ 9
50	-64	-50	- 50	-68	-59	-65	-78	12.5000	3	37	37	2.0	10
-38	-48	-40	-43	-45	-60	- 59	-74	14	7	25	25.8333		
-40	-47	-33	-38	-52	-63	-55	-71	15	5	20	21.3750		
-60	-71	- 59	-60	-70	-74	-77	-88	9	2	80	79.4000		ΜΕΙ ΔΚΔ
-39	-48	-43	-41	-57	-57	-53	-75	14.5000	4	28	21.3750		A R R Read Proved To A Party


C. Floor Plan with X,Y Coordinates

